Journal of Electronic Imaging 12(3), 478–488 (July 2003).

Comparing optimal three-dimensional reconstruction for finite motion and optical flow

Kenichi Kanatani
Okayama University
Department of Information Technology
Okayama 700-8530
Japan
E-mail: kanatani@suri.it.okayama-u.ac.jp

Naoya Ohta
Gunma University
Department of Computer Science
Kiryu, Gunma 376-8515
Japan
E-mail: ohta@cs.gunma-u.ac.jp

Abstract. *We present two linear algorithms for 3-D reconstruction: one is for finite motion; the other is for optical flow. We compare their performance by simulation and real-image experiments, using the same data. The two algorithms are both theoretically optimal, extracting maximum information from the input. We observe that the 3-D reconstruction by the finite motion algorithm is generally more accurate than by the optical flow algorithm if point correspondences are identified with optical flow.* © *2003 SPIE and IS&T.*
[DOI: 10.1117/1.1579018]

1 Introduction

Computing the 3-D structure and motion from an image sequence is one of the most important of computer vision tasks and also one of the research areas that are making rapid progress. A vast literature has already appeared on this subject (see, e.g., Refs. 1 and 2 for the latest developments). There have been two approaches to this problem since 1970s when 3-D reconstruction from images began to attract interest: that based on point correspondences over different views, which we call the finite motion approach; and that based on instantaneous image motion, which we call the optical flow approach. Early literature includes Refs. 3–5 for the former and Refs. 6–8 for the latter. This division has lasted to date because of different mathematical and technical disciplines involved.

Mathematically, the finite motion approach is based on vector calculus of triangulation (or epipolar geometry); the optical flow approach is based on differential calculus of the gray levels (or the gradient constraint). Technically, the former uses template matching for feature correspondence; the latter uses spatiotemporal filtering for optical flow detection. The latter has been closely associated with human perception psychology (see Refs. 7 and 8).

This division is not crucial, however, if optical flow is regarded as point correspondences over consecutive frames; one can run a finite motion algorithm or an optical flow algorithm for 3-D reconstruction. While the optical flow approach is for only a small motion, the finite motion approach can apply to any motion, large or small. Then, is there any benefit to the optical flow approach as far as 3-D reconstruction is concerned? This question has not been fully answered yet, mainly because of the difficulty of comparing the two approaches on a common ground.

Matching feature points for finite motion, whether by template matching or feature tracking, is always a difficult task. Although many constraints, such as the epipolar and trifocal constraints, can be imposed to remove wrong matches, there is no guarantee that the resulting matches are correct. Even if they are correct, their accuracy is usually limited to 1 pixel. Optical flow, on the other hand, could be obtained with subpixel accuracy by spatiotemporal filtering of a long image sequence. One can also estimate the true instantaneous velocity rather than point correspondences over consecutive images.

Evidently, one cannot compare different algorithms if their input is different. This has been the main difficulty for comparing the two approaches. In this paper, we identify optical flow as point correspondence over two images, although this may somewhat impair the advantage of optical flow. On the other hand, various multi-image algorithms are known for finite motion over many images,[1,2] but here we focus on only two-view analysis. This simplification is for concentrating on only 3-D reconstruction, disregarding the quality of input data. Our comparison is restricted in this sense.

In comparing 3-D reconstruction algorithms based on different principles, one cannot make a fair comparison un-

Paper JEI 01059 received Sep. 11, 2001; revised manuscript received Dec. 30, 2002; accepted for publication Jan. 17, 2003.
1017-9909/2003/$15.00 © 2003 SPIE and IS&T.

less both are optimally built on their principles. Thus, we first present an optimal algorithm for finite motion and an optimal algorithm for optical flow. They are based on a statistical model of uncertainty of feature locations. The optimization technique we use is already known but has been published only in fragments.[9–11] Here, we give a self-consistent description of our algorithm for finite motion. For optical flow, the major part was presented in our previously published paper,[12] with which the present paper should be combined for consistent description.

Section 2 introduces the fundamental matrices, which are the basis of 3-D reconstruction for both finite motion and optical flow. Section 3 defines a statistical model of feature uncertainty and gives a theoretical bound on the estimation accuracy. Section 4 presents an optimal algorithm for computing the fundamental matrix for finite motion. It is also confirmed that the solution indeed falls in the vicinity of the theoretical bound. Section 5 describes the 3-D reconstruction procedure for both finite motion and optical flow, and Sec. 6 compares their results. Section 7 presents our conclusions.

2 Fundamental Matrices

2.1 *Definition*

Let $\{(x_\alpha, y_\alpha)\}$ and $\{(x'_\alpha, y'_\alpha)\}$, $\alpha = 1, \ldots, N$, be image coordinates of two sets of N points on two different images; the image coordinate system is defined arbitrarily for each camera. Define vectors

$$\mathbf{x}_\alpha = \begin{pmatrix} x_\alpha / f_0 \\ y_\alpha / f_0 \\ 1 \end{pmatrix}, \quad \mathbf{x}'_\alpha = \begin{pmatrix} x'_\alpha / f_0 \\ y'_\alpha / f_0 \\ 1 \end{pmatrix}, \tag{1}$$

where f_0 is an appropriate scale factor* chosen so that x_α / f_0, y_α / f_0, x'_α / f_0, and y'_α / f_0 have an order 1. For brevity, we call the point having coordinates (x_α, y_α) simply "point $\mathbf{x}_\alpha$."

If $\mathbf{x}_\alpha$ and $\mathbf{x}'_\alpha$ are projections of the same point in the scene, they must satisfy the constraint

$$(\mathbf{x}_\alpha, \mathbf{F}\mathbf{x}'_\alpha) = 0, \tag{2}$$

known as the epipolar equation.[1,2] Here, $\mathbf{F}$ is a matrix of determinant 0, called the fundamental matrix.[1,2] Throughout this paper, we denote by $(\mathbf{a}, \mathbf{b})$ the inner product of vectors $\mathbf{a}$ and $\mathbf{b}$. Since the absolute scale of $\mathbf{F}$ is unconstrained, we normalize it to $\|\mathbf{F}\| = 1$, where the norm of a matrix $\mathbf{A} = (A_{ij})$ is defined by $\|\mathbf{A}\| = (\Sigma_{i,j=1}^{3} A_{ij}^2)^{1/2}$.

If the motion is small, we can write

$$\mathbf{x}'_\alpha = \mathbf{x}_\alpha + \dot{\mathbf{x}}_\alpha \Delta t, \tag{3}$$

to a first approximation, where Δt is the interframe time lapse.‡ The vectors $\{\dot{\mathbf{x}}_\alpha\}$ describing the velocities on the image plane are called the optical flow. Substituting Eq. (3) into Eq. (2) and taking a first approximation, we obtain the following flow epipolar equation[12–14]:

$$(\mathbf{x}_\alpha, \mathbf{W}\dot{\mathbf{x}}_\alpha) + (\mathbf{x}_\alpha, \mathbf{C}\mathbf{x}_\alpha) = 0. \tag{4}$$

Here, $\mathbf{W} = (W_{ij})$ is an antisymmetric matrix, and $\mathbf{C} = (C_{ij})$ is a symmetric matrix. If we define

$$\mathbf{w} = \begin{pmatrix} W_{32} \\ W_{13} \\ W_{21} \end{pmatrix}, \tag{5}$$

the following decomposability condition holds[12–14]:

$$(\mathbf{w}, \mathbf{C}\mathbf{w}) = 0. \tag{6}$$

This constraint results from the rank constraint $\det \mathbf{F} = 0$ for finite motion. The matrices $\mathbf{W}$ and $\mathbf{C}$ are called the flow fundamental matrices.[12]

2.2 *Computation*

Computation of the fundamental matrix $\mathbf{F}$ has been studied by many researchers.[15–23] Common approaches are the bundle-adjustment and the linear algorithm.

The bundle-adjustment is known to be optimal under Gaussian noise, satisfying the Cramer-Rao lower bound (CRLB). However, a parameter space in very high dimensions must be searched, and a good initial guess is required so that the search is not trapped into local minima.[24]

The linear algorithm is based on the observation that the epipolar constraint of Eq. (2) is linear in $\mathbf{F}$. This enables us to compute the solution by simply solving an eigenvalue problem, provided the rank constraint $\det \mathbf{F} = 0$ is ignored.[3,4,16] However, this efficiency sacrifices the accuracy, causing a large statistical bias.[10]

In this paper, we present a modification to the linear algorithm such that the resulting performance is comparable to the bundle-adjustment: we first apply a technique called renormalization,[10] which iteratively removes the statistical bias inherent to the linear algorithm; we then impose the rank constraint in a statistically optimal manner. We also give a (non-CRLB type) theoretical bound on the accuracy of $\mathbf{F}$ by generalizing the uncertainty analysis of Csurka *et al.*[25] We demonstrate that the resulting solution indeed falls in the vicinity of the accuracy bound. The same strategy can also be applied to the computation of the flow fundamental matrices $\mathbf{W}$ and $\mathbf{C}$. An optimal algorithm that delivers results in the vicinity of the theoretical accuracy bound is described in our companion paper.[12]

3 Uncertainty Model and Accuracy Bound

3.1 *Statistical Model of Uncertainty*

We view $\{\mathbf{x}_\alpha\}$ and $\{\mathbf{x}'_\alpha\}$ as perturbed from their true locations $\{\overline{\mathbf{x}}_\alpha\}$ and $\{\overline{\mathbf{x}}'_\alpha\}$ that satisfy the epipolar constraint of Eq. (2). We write

$$\mathbf{x}_\alpha = \overline{\mathbf{x}}_\alpha + \Delta\mathbf{x}_\alpha, \quad \mathbf{x}'_\alpha = \mathbf{x}'_\alpha + \Delta\mathbf{x}'_\alpha, \tag{7}$$

*For example, we can take it to be the size of the image frame.
‡In practice, the interframe laps is taken to be unit time for convenience.

and regard $\Delta\mathbf{x}_\alpha$ and $\Delta\mathbf{x}'_\alpha$ as independent Gaussian random variables of mean $\mathbf{0}$ but not necessarily isotropic or homogeneous. We call $\mathbf{V}[\mathbf{x}_\alpha]=E[\Delta\mathbf{x}_\alpha\Delta\mathbf{x}_\alpha^T]$ and $\mathbf{V}[\mathbf{x}'_\alpha]=E[\Delta\mathbf{x}'_\alpha\Delta\mathbf{x}'^T_\alpha]$ the covariance matrices of $\mathbf{x}_\alpha$ and $\mathbf{x}'_\alpha$, respectively, where $E[\cdot]$ denotes expectation. In practice, we need not know the absolute covariance values; it suffices to know them up to scale. So, we write

$$\mathbf{V}[\mathbf{x}_\alpha]=\epsilon^2\mathbf{V}_0[\mathbf{x}_\alpha],\quad \mathbf{V}[\mathbf{x}'_\alpha]=\epsilon^2\mathbf{V}_0[\mathbf{x}'_\alpha], \tag{8}$$

and assume that $\mathbf{V}_0[\mathbf{x}_\alpha]$ and $\mathbf{V}_0[\mathbf{x}'_\alpha]$, which we call the normalized covariance matrices, are known but the constant ϵ, which we call the noise level, is unknown. The normalized covariance matrices can be estimated from the Hessian of the residual surface of template matching.[26] Since the third components of $\mathbf{x}_\alpha$ and $\mathbf{x}'_\alpha$ are identically 1, the matrices $\mathbf{V}_0[\mathbf{x}_\alpha]$ and $\mathbf{V}_0[\mathbf{x}'_\alpha]$ are singular with third columns and third rows filled with zeros.

If the noise has the same isotropic distribution everywhere, we have

$$\mathbf{V}_0[\mathbf{x}_\alpha]=\mathbf{V}_0[\mathbf{x}'_\alpha]=\mathrm{diag}(1,1,0), \tag{9}$$

where diag $(\cdots)$ denotes the diagonal matrix with diagonal elements $\cdots$. We use Eq. (9) as the default value when no information is available about the noise behavior. It has been experimentally confirmed[26] that Eq. (9) is a good approximation if feature points are extracted using a corner detector (see Refs. 27–31 for existing techniques) or by hand, because corner detectors are so designed and humans are so inclined as to find those points that are distinct from the surroundings in all directions, which means they have covariance matrices more or less in the form of Eq. (9).

If a corresponding pair of $\mathbf{x}_\alpha$ and $\mathbf{x}'_\alpha$ is identified with "optical flow" $\dot{\mathbf{x}}_\alpha=\mathbf{x}'_\alpha-\mathbf{x}_\alpha$ at the "midpoint" $(\mathbf{x}_\alpha+\mathbf{x}'_\alpha)/2$, the flow $\dot{\mathbf{x}}_\alpha$ and the location $\mathbf{x}_\alpha$ are uncorrelated and have their normalized covariance matrices $\mathbf{V}_0[\mathbf{x}_\alpha]+\mathbf{V}_0[\mathbf{x}'_\alpha]$ and $(\mathbf{V}_0[\mathbf{x}_\alpha]+\mathbf{V}_0[\mathbf{x}'_\alpha])/4$, respectively. It follows that the corresponding default noise model for the optical flow approach is given as follows:[12]

$$\mathbf{V}_0[\dot{\mathbf{x}}_\alpha]=2\ \mathrm{diag}\ (1,1,0),\quad \mathbf{V}_0[\mathbf{x}_\alpha]=\tfrac{1}{2}\mathrm{diag}\ (1,1,0). \tag{10}$$

3.2 *Theoretical Accuracy Bound*

Let $\hat{\mathbf{F}}$ be an estimate of the fundamental matrix, and $\bar{\mathbf{F}}$ its true value. The uncertainty of the estimate $\hat{\mathbf{F}}$ is measured by its *covariance tensor*

$$\mathcal{V}[\mathbf{F}]=E[\mathcal{P}((\hat{\mathbf{F}}-\bar{\mathbf{F}})\otimes(\hat{\mathbf{F}}-\bar{\mathbf{F}}))\mathcal{P}^T], \tag{11}$$

where the operator $\otimes$ denotes tensor product: for matrices $\mathbf{A}=(A_{ij})$ and $\mathbf{B}=(B_{ij})$, the $(ijkl)$ element of their tensor product is $A_{ij}B_{kl}$. For tensors $\mathcal{P}=(P_{ijkl})$ and $\mathcal{T}=(T_{ijkl})$, the product $\mathcal{P}\mathcal{T}\mathcal{P}^T$ is a tensor whose $(ijkl)$ element is $\Sigma_{m,n,p,q=1}^3P_{ijmn}P_{klpq}T_{mnpq}$. The $(ijkl)$ element of the tensor $\mathcal{P}=(P_{ijkl})$ in Eq. (11) is given by

$$P_{ijkl}=\delta_{ik}\delta_{jl}-\bar{F}_{ij}\bar{F}_{kl}, \tag{12}$$

where δ_{ij} is the Kronecker delta, taking 1 for $i=j$ and 0 otherwise.

Invoking the general theory of statistical optimization,[10] we can derive a (non-CRLB type) lower bound on the covariance tensor $\mathcal{V}[\hat{\mathbf{F}}]$: If we define the moment tensor $\mathcal{M}=(\bar{M}_{ijkl})$ by

$$\bar{\mathcal{M}}=\frac{1}{N}\sum_{\alpha=1}^{N}\bar{W}_\alpha\bar{\mathbf{x}}_\alpha\otimes\bar{\mathbf{x}}'_\alpha\otimes\bar{\mathbf{x}}_\alpha\otimes\bar{\mathbf{x}}'_\alpha, \tag{13}$$

$$\bar{W}_\alpha=\frac{1}{(\bar{\mathbf{x}}'_\alpha,\bar{\mathbf{F}}^T\mathbf{V}_0[\mathbf{x}_\alpha]\bar{\mathbf{F}}\bar{\mathbf{x}}'_\alpha)+(\bar{\mathbf{x}}_\alpha,\bar{\mathbf{F}}\mathbf{V}_0[\mathbf{x}'_\alpha]\bar{\mathbf{F}}^T\bar{\mathbf{x}}_\alpha)}, \tag{14}$$

the accuracy bound is given in the form

$$\mathcal{V}[\hat{\mathbf{F}}]>\frac{\epsilon^2}{N}(\mathcal{P}^{\mathcal{S}}\bar{\mathcal{M}}\mathcal{P}^{\mathcal{S}T})_7^-, \tag{15}$$

where $\mathcal{T}>\mathcal{S}$ for tensors $\mathcal{T}$ and $\mathcal{S}$ means that $\mathcal{T}-\mathcal{S}$ is a positive semidefinite tensor, and the operation $(\cdot)_r^-$ denotes the (Moore-Penrose) generalized inverse of rank r (discussed later). The $(ijkl)$ element of the tensor $\mathcal{P}^{\mathcal{S}}=(P^{\mathcal{S}}_{ijkl})$ in Eq. (15) is given by

$$P^{\mathcal{S}}_{ijkl}=\delta_{ik}\delta_{jl}-\frac{\bar{F}^\dagger_{ji}\bar{F}^\dagger_{lk}}{\|\bar{\mathbf{F}}^\dagger\|^2}, \tag{16}$$

where $\bar{\mathbf{F}}^\dagger$ is the cofactor matrix of $\bar{\mathbf{F}}$.

For a tensor $\mathcal{T}=(T_{ijkl})$, a matrix $\mathbf{A}=(A_{ij})$, and a scalar λ, we say that $\mathbf{A}$ is an eigenmatrix of $\mathcal{T}$ with eigenvalue λ if $\mathcal{T}\mathbf{A}=\lambda\mathbf{A}$, where the product $\mathcal{T}\mathbf{A}$ is a matrix whose (ij) element is $\Sigma_{k,l=1}^3T_{ijkl}A_{kl}$. The eigenmatrices and eigenvalues of a tensor can be computed by identifying a tensor and a matrix with a 9×9 matrix and a 9-D vector.[10]

A tensor $\mathcal{T}=(T_{ijkl})$ is said to be symmetric if $T_{ijkl}=T_{klij}$. A symmetric $3\times3\times3\times3$ tensor has nine real eigenvalues $\{\lambda_i\}$. The corresponding eigenmatrices $\{\mathbf{U}_i\}$ can be chosen to be an orthogonal system of matrices of unit norm, where the inner product of matrices $\mathbf{A}=(A_{ij})$ and $\mathbf{B}=(B_{ij})$ is defined by $(\mathbf{A};\mathbf{B})=\Sigma_{i,j=1}^3A_{ij}B_{ij}$. A symmetric tensor is positive semidefinite if its eigenvalues are all nonnegative.

Let $\lambda_1\geqslant\cdots\geqslant\lambda_9(\geqslant0)$ be the eigenvalues of a positive semidefinite symmetric tensor $\mathcal{T}$, and let $\{\mathbf{U}_1,\ldots,\mathbf{U}_9\}$ be the corresponding orthonormal set of eigenmatrices of unit norm. If $\lambda_r>0$, the (Moore-Penrose) generalized inverse of $\mathcal{T}$ of rank r is computed as follows:

$$(\mathcal{T})_r^-=\sum_{i=1}^{r}\frac{\mathbf{U}_i\otimes\mathbf{U}_i}{\lambda_i}. \tag{17}$$

The root-mean-square (rms) error of an estimate $\hat{\mathbf{F}}$ is defined by

$$\mathrm{rms}[\hat{\mathbf{F}}]=(E(\|\mathcal{P}(\hat{\mathbf{F}}-\bar{\mathbf{F}})\|^2])^{1/2}. \tag{18}$$

Since $\hat{\mathbf{F}}$ and $\bar{\mathbf{F}}$ are both normalized to unit norm, we have $0 \leqslant \mathrm{rms}[\hat{\mathbf{F}}] \leqslant 1$. From Eq. (15), we have

$$\mathrm{rms}[\hat{\mathbf{F}}] \geqslant \frac{\epsilon}{\sqrt{N}} [\mathrm{tr}(\mathcal{P}^{\mathcal{S}} \bar{\mathcal{M}} \mathcal{P}^{\mathcal{S}T})_7^-]^{1/2}, \tag{19}$$

where the trace tr $\mathcal{T}$ of a tensor $\mathcal{T}=(T_{ijkl})$ is defined by

$$\mathrm{tr}\ \mathcal{T} = \sum_{k,l=1}^{3} T_{klkl}. \tag{20}$$

A similar accuracy bound can also be obtained for the flow fundamental matrices **W** and **C**. The details are described in the companion paper.[12]

4 Optimal Algorithm for the Fundamental Matrix

Our algorithm first computes the fundamental matrix **F** optimally by a technique called renormalization[10] without considering the rank constraint $\det \mathbf{F}=0$ and then imposes the rank constraint $\det \mathbf{F}=0$ in a statistically optimal manner. We show that accuracy is not lost by this two-stage cascading.

4.1 *Renormalization*

The renormalization algorithm proceeds as follows:

1. Let $c=0$, $W_\alpha=1$, $\alpha=1,\ldots,N$, and $J=\infty$, where the symbol ∞ means a very large number, e.g., 10^{10}.
2. Compute the tensors $\mathcal{M}=(M_{ijkl})$ and $\mathcal{N}=(N_{ijkl})$ as follows:

$$M_{ijkl} = \frac{1}{N} \sum_{\alpha=1}^{N} W_\alpha x_{\alpha(i)} x'_{\alpha(j)} x_{\alpha(k)} x'_{\alpha(l)}, \tag{21}$$

$$N_{ijkl} = \frac{1}{N} \sum_{\alpha=1}^{N} W_\alpha (\mathbf{V}_0[\mathbf{x}_\alpha]_{ik} x'_{\alpha(j)} x'_{\alpha(l)} + \mathbf{V}_0[\mathbf{x}'_\alpha]_{jl} x_{\alpha(i)} x_{\alpha(k)}), \tag{22}$$

   where $x_{\alpha(i)}$ and $x'_{\alpha(i)}$ are the i-th components of $\mathbf{x}_\alpha$ and $\mathbf{x}'_\alpha$, respectively, and $\mathbf{V}_0[\mathbf{x}_\alpha]_{ij}$ and $\mathbf{V}_0[\mathbf{x}'_\alpha]_{ij}$ are the (ij) elements of $\mathbf{V}_0[\mathbf{x}_\alpha]_{ij}$ and $\mathbf{V}_0[\mathbf{x}'_\alpha]_{ij}$, respectively.

3. Compute the nine eigenvalues $\lambda_1 \geqslant \ldots \geqslant \lambda_9$ of the tensor $\mathcal{M}$ and the corresponding orthonormal set $\{\mathbf{F}_1,\ldots,\mathbf{F}_9\}$ of eigenmatrices of unit norm.
4. Update c by:

$$c \leftarrow c + \frac{\lambda_9}{(\mathbf{F}_9; \mathcal{N}\mathbf{F}_9)}. \tag{23}$$

5. Compute W_α, $\alpha=1,\ldots,N$, by:

$$W_\alpha = \frac{1}{(\mathbf{x}'_\alpha, \mathbf{F}_9^T \mathbf{V}_0[\mathbf{x}_\alpha] \mathbf{F}_9 \mathbf{x}'_\alpha) + (\mathbf{x}_\alpha, \mathbf{F}_9 \mathbf{V}_0[\mathbf{x}'_\alpha] \mathbf{F}_9^T \mathbf{x}_\alpha)}. \tag{24}$$

6. Compute the tensors $\mathcal{M}$ and $\mathcal{N}$ by Eqs. (21) and (22), and let

$$J' \leftarrow J, \quad J \leftarrow (\mathbf{F}_9; \mathcal{M}\mathbf{F}_9). \tag{25}$$

7. If $J' < J$, let $J \leftarrow J'$. (The variable J stores the minimum residual.) Else, compute the nine eigenvalues $\lambda_1 \geqslant \ldots \geqslant \lambda_9$ of the tensor

$$\hat{\mathcal{M}} = \mathcal{M} - c\mathcal{N}, \tag{26}$$

   and the corresponding orthonormal set $\{\mathbf{F}_1,\ldots,\mathbf{F}_9\}$ of eigenmatrices of unit norm.

8. Repeat steps 4–7 until $J' \leqslant J$ or $|\lambda_9| \approx 0$. This guarantees the iterations will converge.
9. Let **F** take the value $\mathbf{F}_9$, this being our estimate of the fundamental matrix.

The preceding procedure is based on the observation that, in the absence of noise, the fundamental matrix **F** is the eigenmatrix of the moment tensor $\mathcal{M}$ defined in Eq. (21) with eigenvalue 0. It can be shown that, in the presence of noise, $\mathcal{M}$ is statistically biased from its true value to a first approximation by a constant times the tensor $\mathcal{N}$ defined in Eq. (22). By Eq. (26), we iteratively remove the bias in $\mathcal{M}$ in such a way that the smallest eigenvalue of $\mathcal{M}$ converges to zero.

Recently, there has been considerable progress in this type of statistical optimization, and rigorous mathematical analyses and improved techniques have been presented in various forms.[32–35] The corresponding renormalization procedure can also be obtained for the flow fundamental matrices **W** and **C**. The details are described in the companion paper.[12]

4.2 *Optimal Correction*

We next apply a correction to **F**, shifting it iteratively to the nearest value that satisfies the rank constraint $\det \mathbf{F}=0$. This shift is determined optimally with respect to the covariance tensor, which can be evaluated using the eigenvalues and eigenmatrices resulting from the preceding renormalization procedure. The correction steps go as follows.

First, compute the normalized covariance tensor of **F** by:

$$\mathcal{V}_0[\mathbf{F}] = \frac{1}{N} \sum_{i=1}^{8} \frac{\mathbf{F}_i \otimes \mathbf{F}_i}{\lambda_i}. \tag{27}$$

Then, repeat the following computation until $\det \mathbf{F} \approx 0$.

1. Update **F** as follows:

$$\mathbf{F} \leftarrow N\left[\mathbf{F} - \frac{(\det \mathbf{F}) \mathcal{V}_0[\mathbf{F}] \mathbf{F}^{\dagger T}}{(\mathbf{F}^{\dagger T}; \mathcal{V}_0[\mathbf{F}] \mathbf{F}^{\dagger T})}\right]. \tag{28}$$

   The operator $N[\,\cdot\,]$ denotes normalization of the norm to 1: $N[\mathbf{F}] = \mathbf{F}/\|\mathbf{F}\|$.

2. Compute the projection tensor $\mathcal{P}=(P_{ijkl})$ as follows:

$$P_{ijkl} = \delta_{ik}\delta_{jl} - F_{ij}F_{kl}. \tag{29}$$

3. Update the normalized covariance tensor $\mathcal{V}_0[\mathbf{F}]$ as

follows:

$$\mathcal{V}_0[\mathbf{F}]_{ijkl} \leftarrow \sum_{m,n,p,q=1}^{3} P_{ijmn} P_{klpq} \mathcal{V}_0[\mathbf{F}]_{mnpq}. \quad (30)$$

This operation projects the error distribution onto the space orthogonal to $\mathbf{F}$.

It can be shown[10] that the renormalization solution $\mathbf{F}$ has the normalized covariance tensor $\mathcal{V}_0[\hat{\mathbf{F}}]$ given in Eq. (27). The preceding procedure enforces the rank constraint $\det \mathbf{F} = 0$ by iteratively updating $\mathbf{F}$ along the shortest path in the sense of the Mahalanobis distance defined by the normalized covariance tensor $\mathcal{V}_0[\hat{\mathbf{F}}]$. It can be proved that renormalization coupled with this type of correction produces a solution that attains the theoretical accuracy bound in the first order.[10] The corresponding optimal correction procedure is also applied to the flow fundamental matrices $\mathbf{W}$ and $\mathbf{C}$. The details are described in the companion paper.[12]

4.3 *Program Package*

The algorithm described in the preceding is implemented in C++ and placed on our Web page.[§] It outputs a solution $\hat{\mathbf{F}}$ along with its standard deviation pair $\{\mathbf{F}^{(+)}, \mathbf{F}^{(-)}\}$. These are the values in the parameter space that are separated from $\hat{\mathbf{F}}$ by the standard deviation in the two directions along which errors implied by Eq. (15) are the most likely to occur.

We evaluate the right-hand side of Eq. (15) by substituting the data and the estimate $\hat{\mathbf{F}}$ for their true values. The square noise level ϵ^2 in the expression can be estimated by

$$\hat{\epsilon}^2 = \frac{J}{1 - 8/N}, \quad (31)$$

using the value J returned by the renormalization procedure. This type of estimation is known to give a good approximation to the true value.[10]

Let $\lambda_{\max}$ be the maximum eigenvalue of the thus evaluated tensor on the right-hand side of Eq. (15), and let $\mathbf{U}_{\max}$ be the corresponding eigenmatrix of unit norm. The standard deviation pair is defined by

$$\mathbf{F}^{(+)} = N[\hat{\mathbf{F}} + (\lambda_{\max})^{1/2} \mathbf{U}_{\max}],$$

$$\mathbf{F}^{(-)} = N[\hat{\mathbf{F}} - (\lambda_{\max})^{1/2} \mathbf{U}_{\max}]. \quad (32)$$

If $\mathbf{F}^{(+)}$ and $\mathbf{F}^{(-)}$ coincide up to, say, three significant digits, the solution $\hat{\mathbf{F}}$ is likely to have accuracy up to approximately three significant digits.

The fundamental matrix cannot be defined uniquely if the feature points are in a degenerate configuration. This occurs, for example, when the camera translation is zero or all the feature points are on a special quadric called a critical surface, a typical instance of which is a planar surface.[9]

[§]http://www.ail.cs.gunma-u.ac.jp/Labo/programs-e.html

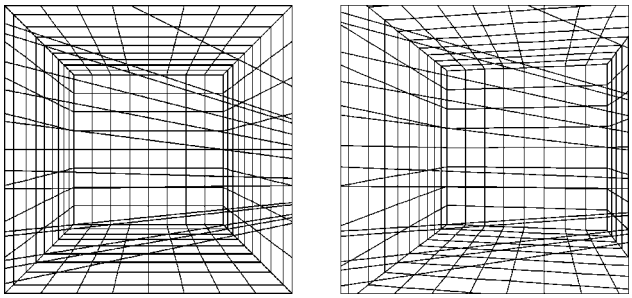


Fig. 1 Simulated images of a 3-D scene and epipolars.

If $\lambda_{\max}$ in Eqs. (32) is predicted to be approximately 1 in the course of computation, our program judges that degeneracy has occurred and stops the computation after issuing a warning message. The corresponding program package for computing the flow fundamental matrices $\mathbf{W}$ and $\mathbf{C}$ is also available at the same site.

4.4 *Simulated Experiments*

Figure 1 shows simulated 512×512-pixel images of a 3-D grid environment. They are supposedly captured by a moving camera with different focal lengths. Some of the epipolars (the images of the lines of sight starting from the projection center of the other camera) are superimposed. Random Gaussian noise of mean 0 and standard deviation σ (pixels) was added to the x and y coordinates of each grid point independently, and the fundamental matrix $\mathbf{F}$ was computed by using the default noise model of Eq. (9). The renormalization converged after three or four iterations.

Figure 2(a) shows a plot of the rms error $(\sum_{a=1}^{100} \|\mathcal{P}(\hat{\mathbf{F}}^a - \bar{\mathbf{F}})\|^2/100)^{1/2}$ over 100 trials for each σ using different noise each time, where $\hat{\mathbf{F}}^a$ is the a-th estimate, $\bar{\mathbf{F}}$ is the true value, and $\mathcal{P}$ is the projection tensor defined in Eq. (12). The symbol $\square$ denotes solutions obtained via the method presented in this paper, and the dotted line indicates the theoretical lower bound derived from Eq. (15). The symbol $\bullet$ denotes renormalization solutions without applying the optimal correction of Sec. 4.2. The symbol $\diamond$ denotes solutions computed by the widely used linear algorithm, often referred to as the least-squares method or the algebraic distance minimization;[16,17] it directly minimizes the sum of the squares of the epipolar constraint of Eq. (2) in the form

$$\frac{1}{N} \sum_{\alpha=1}^{N} (\mathbf{x}_\alpha, \mathbf{F}\mathbf{x}'_\alpha)^2 \rightarrow \min, \quad (33)$$

after the data are normalized as recommended by Hartley.[16]

As we can see from Fig. 2(a), the errors in our estimates practically fall on the theoretical lower bound, which is known to be attained by the bundle adjustment. This confirms that our linear algorithm indeed achieves the accuracy of the bundle adjustment. Figure 2(b) shows the average computation time on a Sun Ultra-30 workstation (Sun OS 5.6). Naturally, our method takes more time than the naive least-squares method, but the theoretical accuracy bound is attained only at this much computation. Similar experi-

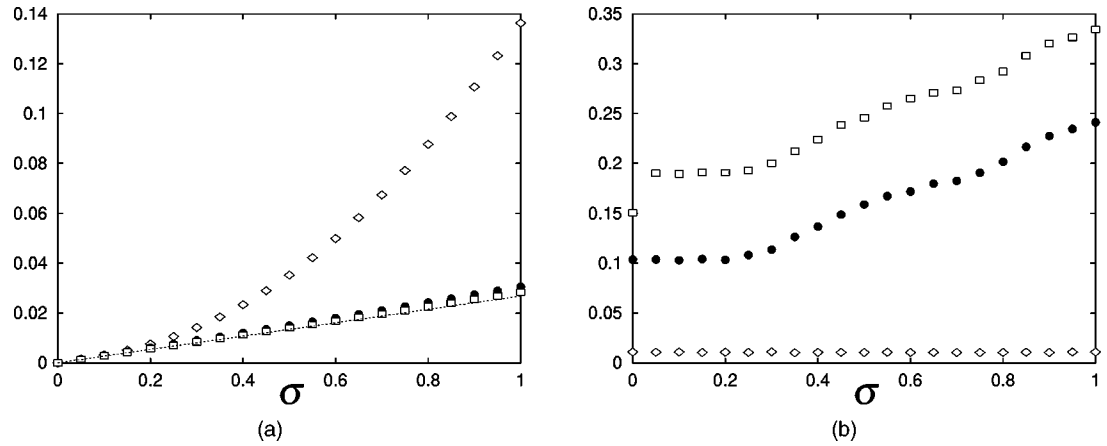


Fig. 2 Accuracy and efficiency of computation: least-squares solutions (◇), solutions without the optimal correction (●), and our solutions (□) for (a) rms error, where the dotted line indicates the theoretical lower bound; and (b) average computation time (in seconds).

ments for confirming the optimality of the flow fundamental matrices **W** and **C** are described in the companion paper.[12]

5 3-D Reconstruction from Two Views

We now describe the 3-D reconstruction procedure for both finite motion and optical flow.

5.1 *Finite Motion Approach*

After the fundamental matrix **F** is computed, the image locations $\mathbf{x}_\alpha$ and $\mathbf{x}'_\alpha$ of each feature point are corrected so as to satisfy the epipolar equation (2) exactly in a statistically optimal way[10]:

$$\hat{\mathbf{x}}_\alpha = \mathbf{x}_\alpha - \frac{E[\mathbf{x}_\alpha, \mathbf{x}'_\alpha]}{V[\mathbf{x}_\alpha, \mathbf{x}'_\alpha]} \mathbf{V}_0[\mathbf{x}_\alpha] \mathbf{F} \mathbf{x}'_\alpha ,$$

$$\hat{\mathbf{x}}'_\alpha = \mathbf{x}'_\alpha - \frac{E[\mathbf{x}_\alpha, \mathbf{x}'_\alpha]}{V[\mathbf{x}_\alpha, \mathbf{x}'_\alpha]} \mathbf{V}_0[\mathbf{x}'_\alpha] \mathbf{F}^T \mathbf{x}_\alpha . \tag{34}$$

Here, we have defined

$$E[\mathbf{x}_\alpha, \mathbf{x}'_\alpha] = (\mathbf{x}_\alpha, \mathbf{F}\mathbf{x}'_\alpha),$$

$$V[\mathbf{x}_\alpha, \mathbf{x}'_\alpha] = (\mathbf{x}'_\alpha, \mathbf{F}^T \mathbf{V}_0[\mathbf{x}_\alpha] \mathbf{F} \mathbf{x}'_\alpha) + (\mathbf{x}_\alpha, \mathbf{F} \mathbf{V}_0[\mathbf{x}'_\alpha] \mathbf{F}^T \mathbf{x}_\alpha). \tag{35}$$

Letting $\mathbf{x}_\alpha \leftarrow \hat{\mathbf{x}}_\alpha$ and $\mathbf{x}'_\alpha \leftarrow \hat{\mathbf{x}}'_\alpha$, we repeat this procedure until the epipolar equation $E[\hat{\mathbf{x}}_\alpha, \hat{\mathbf{x}}'_\alpha] = 0$ is sufficiently satisfied. The convergence has quadratic speed, so one iteration is almost sufficient. This procedure is equivalent to the triangulation of Hartley and Sturm,[36] which requires solving a sixth-degree polynomial, but the preceding form is far more efficient, as pointed out by Torr and Zisseramann.[20]

In reconstructing the 3-D structure from point correspondences over two images taken by two uncalibrated cameras, all information is encoded[1,2] in the fundamental matrix **F**. Since **F** is defined up to scale and constrained to be $\det \mathbf{F} = 0$, it has seven degrees of freedom. The relative motion of the two cameras is specified by a translation vector **t** and a rotation matrix **R**, but the absolute scale of the translation is indeterminate and a 3-D rotation has three degrees of freedom. So, the motion parameters {**t**, **R**} have five degrees of freedom. It follows that only two camera parameters can be recovered.

A practical choice for them is the focal lengths f and f' of the two cameras, since other parameters can be precalibrated and fixed while zooming changes can occur as the camera moves.** Hartley[37] presented an analytic procedure for computing the focal lengths f and f' from the fundamental matrix **F**. The solution is obtained by applying the singular value decomposition (SVD) and solving linear equations in four unknowns. Pan *et al.*[38,39] reduced this problem to solving cubic equations. Newsam *et al.*[40] refined these algorithms into a combination of SVD and linear equations in three unknowns. Kanatani and Matsunaga[11] reduced the problem to solving a quadratic equation in one variable. Bougnoux[41] presented an explicit formula for f in **F**. The degeneracy condition for the solution to be indeterminate has also been analyzed.[11,40]

Among many mathematically equivalent alternatives, the most convenient may be the following modification of the Bougnoux formula[41] given by Kanatani and Matsunaga:[11]

$$f = f_0 \Bigg/ \left[1 + \frac{\|\mathbf{F}k\|^2 - (k, \mathbf{F}\mathbf{F}^T \mathbf{F}k) \|e' \times k\|^2 / (k, \mathbf{F}k)}{\|e' \times k\|^2 \|\mathbf{F}^T k\|^2 - (k, \mathbf{F}k)^2} \right]^{1/2},$$

$$f' = f_0 \Bigg/ \left[1 + \frac{\|\mathbf{F}^T k\|^2 - (k, \mathbf{F}\mathbf{F}^T \mathbf{F}k) \|e \times k\|^2 / (k, \mathbf{F}k)}{\|e \times k\|^2 \|\mathbf{F}k\|^2 - (k, \mathbf{F}k)^2} \right]^{1/2}. \tag{36}$$

Here, we put $k = (0,0,1)^T$. The symbols e and e' denote the unit eigenvectors of $\mathbf{F}^T$ and **F**, respectively, for eigenvalue 0; they represent the epipoles—the image of the projection

**Strictly speaking, the principal point (the intersection of the optical axis with the image plane) may slightly move as zooming changes, but regarding it as a fixed point is known to be a good approximation.

center of the second camera in the first image and the image of that of the first camera in the second image, respectively.[1,2]

After the focal lengths f and f' have been obtained, we recompute $\{\hat{\mathbf{x}}_\alpha\}$ and $\{\hat{\mathbf{x}}'_\alpha\}$ by replacing f_0 in the vector representation of Eqs. (1) by the corresponding values f and f'. This can be done as follows:

$$\hat{\mathbf{x}}_\alpha \leftarrow \mathrm{diag}\left(\frac{f_0}{f}, \frac{f_0}{f}, 1\right)\hat{\mathbf{x}}_\alpha, \quad \mathbf{x}'_\alpha \leftarrow \mathrm{diag}\left(\frac{f_0}{f'}, \frac{f_0}{f'}, 1\right)\mathbf{x}'_\alpha. \tag{37}$$

The motion parameters $\{\mathbf{t},\mathbf{R}\}$ are analytically computed by the following procedure,[9] where the indeterminate scale of the translation $\mathbf{t}$ is normalized to $\|\mathbf{t}\|=1$.

1. Compute the following essential matrix:

$$\mathbf{E} = \mathrm{diag}\left(1,1,\frac{f_0}{f}\right)\mathbf{F}\,\mathrm{diag}\left(1,1,\frac{f_0}{f'}\right). \tag{38}$$

   This removes the dependence of the fundamental matrix on the focal lengths.[11]

2. Compute the unit eigenvector $\mathbf{t}$ of $\mathbf{E}\mathbf{E}^T$ for the smallest eigenvalue. The sign of $\mathbf{t}$ is chosen in such a way that

$$\sum_{\alpha=1}^{N} |\mathbf{t}, \hat{\mathbf{x}}_\alpha, \mathbf{E}\mathbf{x}'_\alpha| > 0. \tag{39}$$

   This is the constraint that the depths of the feature points have the same sign before and after the camera motion.[9]

3. Apply SVD to $-\mathbf{t}\times\mathbf{E}$ as follows:

$$-\mathbf{t}\times\mathbf{E} = \mathbf{V}\mathbf{\Lambda}\mathbf{U}^T. \tag{40}$$

   For a vector $\mathbf{a}$ and a matrix $\mathbf{A}$, we define $\mathbf{a}\times\mathbf{A}$ to be the matrix consisting of columns that are the vector products of $\mathbf{a}$ and the individual columns of $\mathbf{A}$. Also $\mathbf{\Lambda}$ is a diagonal matrix with diagonal elements (singular values) in nonincreasing order, and $\mathbf{V}$ and $\mathbf{U}$ are orthogonal matrices.

4. Compute the rotation $\mathbf{R}$ as follows:

$$\mathbf{R} = \mathbf{V}\,\mathrm{diag}\,(1,1,\det\mathbf{V}\mathbf{U}^T)\mathbf{U}^T. \tag{41}$$

This procedure produces a least-squares solution for $\{\mathbf{t},\mathbf{R}\}$ even if the rank constraint $\det\mathbf{F}=0$ is not strictly satisfied.[9]

The 3-D position of the αth point is given by

$$\hat{\mathbf{r}}_\alpha = \hat{Z}_\alpha\hat{\mathbf{x}}_\alpha, \quad \hat{\mathbf{r}}'_\alpha = \hat{Z}'_\alpha\hat{\mathbf{x}}'_\alpha, \tag{42}$$

with respect to the first and the second camera coordinate systems, respectively, where the depths $\hat{Z}_\alpha$ and $\hat{Z}'_\alpha$ are given as follows[9]:

$$\hat{Z}_\alpha = (\mathbf{t}\times\mathbf{R}\hat{\mathbf{x}}'_\alpha, \mathbf{n}_\alpha), \quad \hat{Z}'_\alpha = (\mathbf{t}\times\hat{\mathbf{x}}_\alpha, \mathbf{n}_\alpha). \tag{43}$$

Here, we have defined

$$\mathbf{n}_\alpha = \frac{\hat{\mathbf{x}}_\alpha\times\mathbf{R}\hat{\mathbf{x}}'_\alpha}{\|\hat{\mathbf{x}}_\alpha\times\mathbf{R}\hat{\mathbf{x}}'_\alpha\|^2}. \tag{44}$$

Finally, we must adjust the signs of the depths because the sign of the fundamental matrix $\mathbf{F}$ is indeterminate. The signs of $\{\hat{Z}_\alpha\}$ and $\{\hat{Z}'_\alpha\}$ are inverted, if necessary, so that

$$\sum_{\alpha=1}^{N} (\mathrm{sgn}[\hat{Z}_\alpha] + \mathrm{sgn}[\hat{Z}'_\alpha]) > 0, \tag{45}$$

where $\mathrm{sgn}[\cdot]$ is the signature function that takes 1, 0, and -1 for $x>0$, $x=0$, and $x<0$, respectively. This operation is necessary, because we may not select the correct solution if we simply compute $\Sigma_{\alpha=1}^{N}(\hat{Z}_\alpha+\hat{Z}'_\alpha)$; a very large positive depth may turn out to be close to $-\infty$ due to noise.

5.2 *Optical Flow Approach*

The 3-D reconstruction from optical flow goes similarly. After optimally computing the flow fundamental matrix $\mathbf{W}$ and $\mathbf{C}$ using the method described in the companion paper,[12] the flow $\dot{\mathbf{x}}_\alpha$ and image location $\mathbf{x}_\alpha$ of each feature point are corrected so as to satisfy the flow epipolar equation (4) exactly in a statistically optimal way[10]:

$$\hat{\dot{\mathbf{x}}}_\alpha = \dot{\mathbf{x}}_\alpha + \frac{E[\dot{\mathbf{x}}_\alpha,\mathbf{x}_\alpha]}{V[\dot{\mathbf{x}}_\alpha,\mathbf{x}_\alpha]}\mathbf{V}_0[\dot{\mathbf{x}}_\alpha]\mathbf{W}\mathbf{x}_\alpha,$$

$$\hat{\mathbf{x}}_\alpha = \mathbf{x}_\alpha - \frac{E[\dot{\mathbf{x}}_\alpha,\mathbf{x}_\alpha]}{V[\dot{\mathbf{x}}_\alpha,\mathbf{x}_\alpha]}\mathbf{V}_0[\mathbf{x}_\alpha](\mathbf{W}\dot{\mathbf{x}}_\alpha + 2\mathbf{C}\mathbf{x}_\alpha). \tag{46}$$

Here, we have defined

$$E[\dot{\mathbf{x}}_\alpha,\mathbf{x}_\alpha] = (\mathbf{x}_\alpha,\mathbf{W}\dot{\mathbf{x}}_\alpha) + (\mathbf{x}_\alpha,\mathbf{C}\mathbf{x}_\alpha),$$

$$V[\dot{\mathbf{x}}_\alpha,\mathbf{x}_\alpha] = (\mathbf{W}\mathbf{x}_\alpha,\mathbf{V}_0[\dot{\mathbf{x}}_\alpha]\mathbf{W}\mathbf{x}_\alpha) + (\mathbf{W}\dot{\mathbf{x}}_\alpha + 2\mathbf{C}\mathbf{x}_\alpha,\mathbf{V}_0[\mathbf{x}_\alpha](\mathbf{W}\dot{\mathbf{x}}_\alpha + 2\mathbf{C}\mathbf{x}_\alpha)). \tag{47}$$

Letting $\dot{\mathbf{x}}_\alpha \leftarrow \hat{\dot{\mathbf{x}}}_\alpha$ and $\mathbf{x}_\alpha \leftarrow \hat{\mathbf{x}}_\alpha$, we repeat this procedure until the flow epipolar equation $E[\hat{\mathbf{x}}_\alpha,\hat{\mathbf{x}}_\alpha]=0$ is sufficiently satisfied. As in the finite motion approach, the convergence has quadratic speed, so one iteration is almost sufficient.

In reconstructing the 3-D structure from optical flow taken by an uncalibrated camera, all information is encoded[12–14] in the flow fundamental matrices $\mathbf{W}$ and $\mathbf{C}$. They are determined up to scale and constrained by the decomposability condition of Eq. (6). Since $\mathbf{W}$ and $\mathbf{C}$ are antisymmetric and symmetric matrices, respectively, they have seven degrees of freedom in total. The instantaneous motion of the camera is specified by the translation velocity $\mathbf{v}$ and the rotation velocity $\boldsymbol{\omega}$, but the absolute scale of the translational motion is indeterminate. So, the motion parameters $\{\mathbf{v},\boldsymbol{\omega}\}$ have five degrees of freedom. As in the case of finite motion, only two camera parameters can be recovered.

A practical choice for them is, as before, the focal length f and its change rate $\dot{f}$. Brooks *et al.*[13] presented a complicated procedure for computing f, $\dot{f}$, $\mathbf{v}$, and $\boldsymbol{\omega}$ from the flow fundamental matrices $\mathbf{W}=(W_{ij})$ and $\mathbf{C}=(C_{ij})$. Here, we

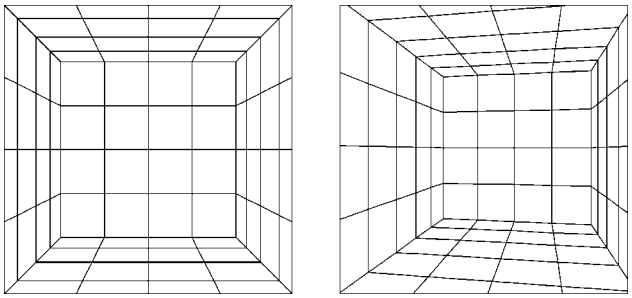


Fig. 3 Simulated images of a 3-D scene.

present an elegant group-theoretical procedure using[42] irreducible representations of the group of 2-D rotations $SO(2)$. The indeterminate scale of the translation velocity $\mathbf{v}$ is normalized to $\|\mathbf{v}\|=1$.

Let w_i be the i'th component of the vector $\mathbf{w}$ defined in Eq. (5), and do the following computation:

$$A=C_{11}+C_{22},\quad \tilde{B}=(C_{11}-C_{22})+2iC_{12},$$

$$\tilde{C}=2(C_{13}+iC_{23}),\quad D=C_{33}, \tag{48}$$

$$\tilde{w}=w_1+iw_2,\quad \tilde{w}'=\frac{\tilde{B}}{\tilde{w}},\quad \omega'_1=\Re[\tilde{\omega}'],$$

$$\omega'_2=\Im[\tilde{\omega}'],\quad \omega_3=-\frac{A+(\tilde{w},\tilde{\omega}')}{2w_3}, \tag{49}$$

$$f'=\left(-\frac{D}{(\tilde{w},\tilde{\omega}')}\right)^{1/2},\quad \tilde{\phi}=\frac{\tilde{C}-f'^2w_3\tilde{\omega}'}{\tilde{w}},$$

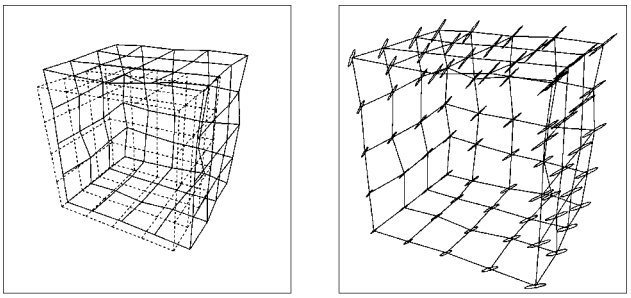
$$\omega_3=\Re[\tilde{\phi}],\quad \dot{f}'=-f'\Im[\tilde{\phi}], \tag{50}$$

$$\omega_1=f'\omega'_1,\quad \omega_2=f'\omega'_2,\quad f=f'f_0,\quad \dot{f}=\dot{f}'f_0,$$

$$\mathbf{v}=N\left[\begin{pmatrix} w_1 \\ w_2 \\ (f/f_0)w_3 \end{pmatrix}\right]. \tag{51}$$

Here, i is the imaginary unit. The quantities with tildes are complex numbers: $\Re[\cdot]$ and $\Im[\cdot]$ denote the real and imaginary parts, respectively. We define the "inner product" of complex numbers $z=x+iy$ and $z'=x'+iy'$ by $(z,z')=xx'+yy'$. The operation $N[\cdot]$ designates normalization into a unit vector: $N[\mathbf{a}]=\mathbf{a}/\|\mathbf{a}\|$. Note that ω_3 is computed in two ways by the fifth of Eqs. (49) and the third of Eqs. (50). The decomposability condition of Eq. (6) requires that the two values coincide.

The 3-D positions of the feature points are reconstructed as follows. First, we recompute $\{\hat{\mathbf{x}}_\alpha\}$ and $\{\hat{\mathbf{x}}_\alpha\}$ by replacing f_0 by its true value f and incorporating its change rate $\dot{f}$. This is done as follows:



(a) (b)

Fig. 4 Finite motion approach: (a) reconstructed shape (solid lines) and the true shape (dotted lines); and (b) uncertainty ellipsoids of the grid points.

$$\hat{\mathbf{x}}_\alpha\leftarrow\frac{f_0}{f}\left(\hat{\mathbf{x}}_\alpha-\frac{\dot{f}}{f}\mathrm{diag}\,(1,1,0)\hat{\mathbf{x}}_\alpha\right),$$

$$\hat{\mathbf{x}}_\alpha\leftarrow\mathrm{diag}\left(\frac{f_0}{f},\frac{f_0}{f},1\right)\hat{\mathbf{x}}_\alpha. \tag{52}$$

The 3-D position of the α'th point is given by

$$\hat{\mathbf{r}}_\alpha=\hat{Z}_\alpha\hat{\mathbf{x}}_\alpha, \tag{53}$$

where the depth $\hat{Z}_\alpha$ is given as follows:[9]

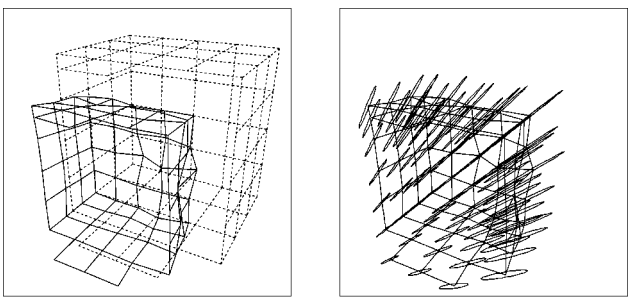
$$\hat{Z}_\alpha=-\frac{(\mathbf{v},\mathbf{S}_\alpha\mathbf{v})}{(\mathbf{v},\mathbf{S}_\alpha(\dot{\mathbf{x}}_\alpha+\boldsymbol{\omega}\times\mathbf{x}_\alpha))}. \tag{54}$$

Here, we have defined

$$\mathbf{S}_\alpha=(\mathbf{I}-\mathbf{x}_\alpha k^T)^T(\mathbf{I}-\mathbf{x}_\alpha k^T), \tag{55}$$

where $k=(0,0,1)^T$. In view of the sign indeterminacy of the flow fundamental matrices $\mathbf{W}$ and $\mathbf{C}$, the signs of $\{\hat{Z}_\alpha\}$ are inverted, if necessary, so that the following condition is satisfied for the same reason as in the finite motion approach:

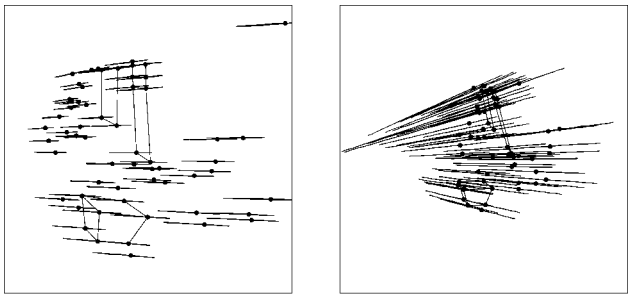
$$\sum_{\alpha=1}^{N}\mathrm{sgn}\,[\hat{Z}_\alpha]>0. \tag{56}$$



(a) (b)

Fig. 5 Optical flow approach: (a) reconstructed shape (solid lines) and the true shape (dotted lines); and (b) uncertainty ellipsoids of the grid points.

Fig. 6 Real images of an indoor scene.



(a) (b)

Fig. 7 Uncertainty ellipsoids of points reconstructed (a) by the finite motion approach and (b) by the optical flow approach.

6 Performance Comparison

We now compare the performance of the finite motion algorithm and the optical flow algorithm by simulation and real-image experiments using the same data.

6.1 *Simulation Experiments*

Figure 3 shows simulated 512×512-pixel images of a 3-D scene. We added independent random Gaussian noise of mean 0 and standard deviation 3 (pixels) to each of the x and y coordinates of the grid points and reconstructed the 3-D shape using the default noise model of Eq. (9).

Figure 4(a) shows the 3-D shape (solid lines) reconstructed by the finite motion algorithm superimposed on the true shape (dotted lines) rescaled to $\|t\|=1$. Figure 4(b) shows uncertainty ellipsoids centered on the reconstructed vertices. We can evaluate the uncertainty of the computed fundamental matrix $\mathbf{F}$ in the form of the covariance tensor so that it can be propagated to the uncertainty of the 3-D reconstruction (we omit the details). The ellipsoids in Fig. 4(b) indicate three times the standard deviation in each orientation. As we can observe, they are very thin needle-like shapes with their major axes approximately in the depth orientation. They are larger for points farther away from the cameras.[††]

Figure 5 shows the corresponding 3-D shape reconstructed by the optical flow algorithm using the default noise model of Eqs. (10). Here, the random Gaussian noise is reduced to standard deviation 0.5 (pixel), because adding more noise would deteriorate the results intolerably. From this, we can observe the poor performance of the optical flow approach even in this low noise level.

6.2 *Real Image Experiments*

Figure 6 shows a pair of real images (512×768 pixels) of an indoor scene. We manually selected feature points as marked in the images and reconstructed the 3-D shape. Figure 7 shows the 3-D shape computed in two ways: Fig. 7(a) with the finite motion approach using the noise model of Eq. (9); and Fig. 7(b) with the optical flow approach using the noise model of Eqs. (10). Wireframes consisting of some of the reconstructed points are shown for visual aide.

On each reconstructed point is centered the uncertainty ellipsoid that indicates the standard deviation in each orientation (this time not magnified). They are like thin needles, showing that the uncertainty is very large along the depth orientation. Although the reconstructed shape itself looks natural for both Figs. 7(a) and 7(b), we can clearly see that the optical flow solution has far larger uncertainty than the finite motion solution.

Figure 8 shows real images (512×768 pixels) of a car. We manually selected feature points as marked in the images and reconstructed the 3-D shape. Figure 9 shows some new views generated by creating a wireframe model from the reconstructed points and mapping the texture to it: the upper row is obtained by the finite motion approach with the noise model of Eq. (9); the lower row is obtained by the optical flow approach with the noise model of Eqs. (10).

Although it is difficult to grasp the exact 3-D shape from static views, we are given a fairly realistic impression of the 3-D shape by continuously changing the viewpoint. After careful observations, however, we find that the 3-D shape is unnaturally deformed in the part far away from the viewer as compared with the front part, which is fairy accurate. We also find that the deformation is larger for the optical flow solution than the finite motion solution.

7 Conclusions

We have presented two linear algorithms for 3-D reconstruction: one is for finite motion; and the other is for optical flow. Both are theoretically optimal in the sense that they extract maximum information from the input. They first compute the fundamental matrix and the flow fundamental matrices by renormalization followed by optimal correction. For each approach, experiments have shown that the solution falls in the vicinity of the theoretical accuracy bound.

We then presented for each approach a 3-D reconstruction procedure based on the computed fundamental matrices. We compared the performance of the two algorithms

Fig. 8 Real images of a car.

[††]The uncertainty shown here is relative to the first camera coordinate system with the camera translation normalized to unit length, so it does not have an absolute meaning since the first camera coordinate system and the camera translation also have their uncertainty. To extract an absolute meaning, we require the gauge theory of uncertainty description.[43]

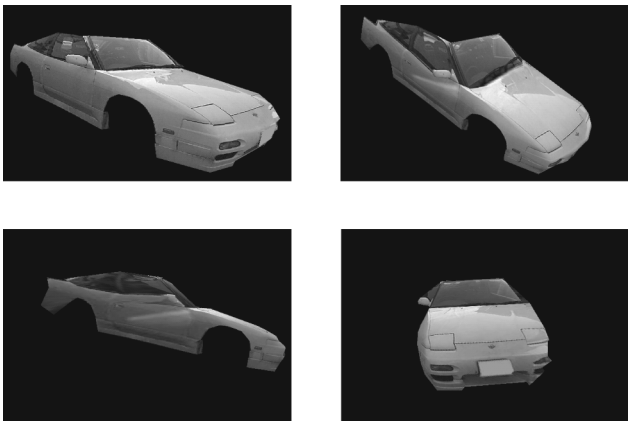


Fig. 9 Reconstructed and texture-mapped 3-D shape using the finite motion approach (above) and the optical flow approach (below).

by simulation and real-image experiments using the same data and observed that the finite motion solution is always superior to the optical flow solution.

Since optical flow is a first-order approximation of finite motion, the computation could be stabler for optical flow when the disparity is very small. We tested this by simulation. Evidently, the 3-D information cannot be obtained if the disparity between the two images is too small whatever method is used. Gradually reducing the disparity, we ran the two algorithms for the same data and found that the optical flow algorithm always collapsed first. In all the experiments we performed, we were unable to find any advantage of the optical flow algorithm in accuracy, efficiency, or stability as far as 3-D reconstruction is concerned.

The reason why so many studies of 3-D reconstruction from optical flow have been done in the past lies perhaps in the ease of its computation and the rich information it contains. Indeed, humans can easily perceive the 3-D structure of the scene by simply looking at its motion. Thus, it is natural that people have sought algorithms for 3-D reconstruction from optical flow. In fact, many authors assert that studying 3-D reconstruction algorithms from optical flow can lead to understand the workings of the human brain.[5,8] Optical flow is also useful for motion segmentation.

For 3-D reconstruction, however, one should use the finite motion algorithm; nothing is gained by using first-order approximations. Of course, our comparison experiments are limited in that point correspondence over two images is identified with optical flow. Establishing correspondence for finite motion by template matching is a difficult task, and its accuracy is limited, while optical flow can be obtained with subpixel accuracy by spatiotemporal filtering of a long image sequence. One can also estimate the true instantaneous velocity rather than point correspondences over consecutive images. We also restricted our investigation to two-view analysis and did not consider the multiframe approach for finite motion.[1,2]

These limitations are an unavoidable consequence of comparing algorithms for different types of data. Evidently, high-accuracy methods do not perform well for poor data, while naive methods may work well for good data. This paper focused on only 3-D reconstruction accuracy, disregarding the quality of input data. The optimality of our procedure is of only theoretical nature based on idealizing assumptions, and the accuracy was tested by only a limited number of experiments. Thus, there is much room for improving the performance for both approaches.

Acknowledgments

The authors thank Hitoshi Mishima of 3D, Inc., Japan, and Yoshiyuki Shimizu of Sharp, Ltd., Japan, for doing various experiments for us. They also thank Chikara Matsunaga of FOR-A, Co. Ltd., Japan, Toshio Ueshiba of AIST, Japan, Long Quan and Bill Triggs of INRIA Rhône Alpes, France, Mike Brooks and his colleagues of the University of Adelaide, Australia, Du Huynh of Murdoch University, Australia, Luis Baumela of Madrid Technical University, Spain, and Lourdes de Agapito of the University of Oxford, United Kingdom for helpful discussions. This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under a Grant in Aid for Scientific Research C(2) (No. 11500113).

References

1. O. Faugeras and Q.-T. Luong, *The Geometry of Multiple Images*, MIT Press, Cambridge, MA, (2001).
2. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK (2000).
3. H. C. Longuet-Higgins, ''A computer algorithm for reconstructing a scene from two projections,'' *Nature (London)* **293**(10), 133–135 (1981).
4. R. Y. Tsai and T. S. Huang, ''Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **6**(1), 12–27 (1984).
5. S. Ullman, *The Interpretation of Visual Motion*, MIT Press, Cambridge, MA (1979).
6. A. R. Bruss and B. K. P. Horn, ''Passive navigation,'' *Comput. Vis. Graph. Image Process.* **21**(1), 3–20 (1983).
7. J. J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston (1979).
8. H. C. Longuet-Higgins and K. Prazdny, ''The interaction of a moving retinal images,'' *Proc. R. Soc. London, Ser. B* **208**, 385–397 (1980).
9. K. Kanatani, *Geometric Computation for Machine Vision*, Oxford University Press, Oxford, UK (1993).
10. K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, Amsterdam (1996).
11. K. Kanatani and C. Matsunaga, ''Closed-form expression for focal lengths from the fundamental matrix,'' in *Proc. 4th Asian Conf. Comput. Vision*, pp. 128–133, Taipei, Taiwan (2000).
12. K. Kanatani, Y. Shimizu, N. Ohta, M. J. Brooks, W. Chojnacki, and A. van den Hengel, ''Fundamental matrix from optical flow: optimal computation and reliability evaluation,'' *J. Electron. Imaging* **9**(2), 194–202 (2000).
13. M. J. Brooks, W. Chojnacki, and L. Baumela, ''Determining the egomotion of an uncalibrated camera from instantaneous optical flow,'' *J. Opt. Soc. Am. A* **14**(10), 2670–2677 (1997).
14. T. Viéville and O. D. Faugeras, ''The first order expansion of motion equations in the uncalibrated case,'' *Comput. Vis. Image Underst.* **64**(1), 128–146 (1996).
15. M. Bober, N. Geogis, and J. Kittler, ''On accurate and robust estimation of fundamental matrix,'' *Comput. Vis. Image Underst.* **72**(1), 39–53 (1998).
16. R. I. Hartley, ''In defense of the eight-point algorithm,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **97**(6), 580–593 (1997).
17. R. I. Hartley, ''Minimizing algebraic error,'' *Philos. Trans. R. Soc. London, Ser. A* **3569**, 1175–1192 (1998).
18. Q.-T. Luong and O. D. Faugeras, ''The fundamental matrix: theory, algorithm, and stability analysis,'' *Int. J. Comput. Vis.* **17**(3), 43–47 (1996).
19. M. Mühlich and R. Mester, ''The role of total least squares in motion analysis,'' in *Proc. 5th Euro. Conf. Comput. Vision*, Lecture Notes in Computer Science, Vol. 1407, pp. 305–321, Springer, Berlin (1998).
20. P. H. S. Torr and A. Zissermann, ''Performance characterization of fundamental matrix estimation under image degradation,'' *Mach. Vision Appl.* **9**, 321–333 (1997).
21. P. H. S. Torr and A. Zisserman, ''Robust detection of degenerate configurations while estimating the fundamental matrix,'' *Comput. Vis. Image Underst.* **71**(3), 312–333 (1998).
22. Z. Zhang, ''Determining the epipolar geometry and its uncertainty: a review,'' *Int. J. Comput. Vis.* **27**(2), 161–195 (1998).

23. Z. Zhang, ''On the optimization criteria used in two-view motion analysis,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(7), 717–729 (1998).
24. B. Triggs, P. F. Mclauchlan, R. I. Hartley, and A. W. Fitzgibbon, ''Bundle adjustment—a modern synthesis,'' in *Vision Algorithms: Theory and Practice*, B. Triggs, A. Zisserman and R. Szeliski, Eds., pp. 298–375, Springer, Berlin (2000).
25. G. Csurka, C. Zeller, Z. Zhang, and O. D. Faugeras, ''Characterizing the uncertainty of the fundamental matrix,'' *Comput. Vis. Image Underst.* **68**(1), 18–36 (1997).
26. Y. Kanazawa and K. Kanatani, ''Do we really have to consider covariance matrices for image features?'' in *Proc. 8th Int. Conf. Comput. Vision*, Vol. 2, pp. 586–591, IEEE Computer Society, Los Alamitos, CA (2001).
27. F. Chabat, G. Z. Yang, and D. M. Hansell, ''A corner orientation detector,'' *Image Vis. Comput.* **17**(10), 761–769 (1999).
28. C. Harris and M. Stephens, ''A combined corner and edge detector,'' in *Proc. 4th Alvey Vision Conf.* pp. 147–151, Manchester, UK (1988).
29. D. Reisfeld, H. Wolfson, and Y. Yeshurun, ''Context-free attentional operators: the generalized symmetry transform,'' *Int. J. Comput. Vis.* **14**(2), 119–130 (1995).
30. C. Schmid, R. Mohr, and C. Bauckhage, ''Evaluation of interest point detectors,'' *Int. J. Comput. Vis.* **37**(2), 151–172 (2000).
31. S. M. Smith and J. M. Brady, ''SUSAN—a new approach to low level image processing,'' *Int. J. Comput. Vis.* **23**(1), 45–78 (1997).
32. W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley, ''On the fitting of surfaces to data with covariances,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1294–1303 (2000).
33. W. Chojnacki, M. J. Brooks, and A. van den Hengel, ''Rationalising the renormalisation method of Kanatani,'' *J. Math. Imaging Vision* **14**(1), 21–38 (2001).
34. Y. Leedan and P. Meer, ''Heteroscedastic regression in computer vision: problems with bilinear constraint,'' *Int. J. Comput. Vis.* **37**(2), 127–150 (2000).
35. B. Matei and P. Meer, ''Reduction of bias in maximum likelihood ellipse fitting,'' in *Proc. 15th Int. Conf. Pattern Recog.*, Vol. 3, pp. 801–806, IEEE Computer Society, Los Alamitos, CA (2000).
36. R. I. Hartley and P. Sturm, ''Triangulation,'' *Computer. Vis. Image Underst.* **68**(2), 146–157 (1997).
37. R. I. Hartley, ''Estimation of relative camera position for uncalibrated cameras,'' in *Proc. 2nd Eur. Conf. Comput. Vision*, Lecture NOtes in Computer Science, Vol. 558, pp. 579–587, Springer, Berlin (1992).
38. H.-P. Pan, M. J. Brooks, and G. Newsam, ''Image resituation: initial theory,'' in *Videometrics IV, Proc. SPIE* **2598**, 162–173 (1995).
39. H.-P. Pan, D. Q. Huynh, and G. Hamlyn, ''Two-image resituation: practical algorithm,'' in *Videometrics IV, Proc. SPIE* **2598**, 174–190 (1995).
40. G. N. Newsam, D. Q. Huynh, M. J. Brooks, and H.-P. Pan, ''Recovering unknown focal lengths in self-calibration: an essentially linear algorithm and degenerate configurations,'' *Int. Arch. Photogram. Remote Sensing* **31**(B3-III), 575–580 (July 1996).
41. S. Bougnoux, ''From projective to Euclidean space under any practical situation, a criticism of self calibration,'' in *Proc. 6th Int. Conf. Comput. Vision.*, pp. 790–796, Narosa, New Delhi (1998).
42. K. Kanatani, *Group-Theoretical Methods in Image Understanding*, Springer, Berlin (1990).
43. K. Kanatani and D. D. Morris, ''Gauges and gauge transformations for uncertainty description of geometric structure with indeterminacy,'' *IEEE Trans. Inf. Theory* **47**(5), 2017–2028 (2001).

Kenichi Kanatani received his PhD degree in applied mathematics from the University of Tokyo in 1979. He was a professor of computer science with Gunma University, Japan, and is currently a professor of information technology with Okayama University, Japan. He is the author of *Group-Theoretical Methods in Image Understanding* (Springer, 1990), *Geometric Computation for Machine Vision* (Oxford, 1993), and *Statistical Optimization for Geometric Computation: Theory and Practice* (Elsevier, 1996). He is an IEEE Fellow.

Naoya Ohta received his ME degree in information science from the Tokyo Institute of Technology in 1985 and his PhD degree in applied mathematics from the University of Tokyo in 1998. He engaged in research and development of image processing systems with the Pattern Recognition Research Laboratories of NEC, Japan, and was a research affiliate with the Media Laboratory at the Massachusetts Institute of Technology (MIT) from 1991 to 1992. He is currently an associate professor of computer science with Gunma University, Japan. His research interests include image processing and computer vision.